

SV3D: Novel Multi-view Synthesis and 3D Generation from a Single Image using Latent Video Diffusion

Year: 2024 | Citations: 298 | Authors: Vikram S. Voleti, Chun-Han Yao, Mark Boss, Adam Letts, David Pankratz

Abstract

We present Stable Video 3D (SV3D) -- a latent video diffusion model for high-resolution, image-to-multi-view generation of orbital videos around a 3D object. Recent work on 3D generation propose techniques to adapt 2D generative models for novel view synthesis (NVS) and 3D optimization. However, these methods have several disadvantages due to either limited views or inconsistent NVS, thereby affecting the performance of 3D object generation. In this work, we propose SV3D that adapts image-to-video diffusion model for novel multi-view synthesis and 3D generation, thereby leveraging the generalization and multi-view consistency of the video models, while further adding explicit camera control for NVS. We also propose improved 3D optimization techniques to use SV3D and its NVS outputs for image-to-3D generation. Extensive experimental results on multiple datasets with 2D and 3D metrics as well as user study demonstrate SV3D's state-of-the-art performance on NVS as well as 3D reconstruction compared to prior works.